Semantic Enabling & Exploitation (SEE) for Assured, Improvisational Workflows

Preliminary Ideas

DAML PI Meeting (18 Oct. 2003)

Brian Kettler, Ph.D.
ISX Corporation
bkettler@isx.com

UNCLASSIFIED

Agenda

- SEE Seedling Goals
- Operational Motivation – problem and examples
- Technology Investment Model
- Applicable Technologies
- **Help Wanted**

- A couple of caveats:
  - ideas are still evolving
  - these are not the final program brief slides
  - words on the slides are my spin (e.g., not necessarily DARPA's or CECOM's)

SEE Seedling Work

- U.Maryland & Dartmouth work (Prof. Jim Hendler & Prof. George Cybenko)
  - technical studies on semantic interoperability
- ISX & U.Maryland (Prof. Jim Hendler & MINDSWAP lab) seedling (4Q FY03 – present)
  - work with DARPA/IXO (Dr. Mark Greaves and Dr. Robert Tenney) to refine operational and technical concepts for a potential new DARPA program

Goal is to articulate an operational and technical vision and "business case" for a new focus on semantic enabling and exploitation for improved interoperability.

The Problem

- In today's world of new missions and partners, improvisational workflows in the field are needed to help a commander meet new information requirements
  - too expensive to design for all possible requirements, even if they were known
- Current (rapid) interoperability efforts are ad hoc, error prone and resource-intensive
  - no time to do extensive design work, semantics are hidden (especially in legacy systems), requires smart programmers to uncover hidden semantics, programming resources limited in the field, etc.
- Interoperability errors have serious operational impact
- This is only going to get worse: increasing operational innovation and tempo require interoperability on the fly in the field

Need: Assured, Improvisational Workflows via Semantic Interoperability.

The Interoperability/Integration Problem

- Some examples - AFSAB study on Database Migration (Interoperability) (2001)
  - AF/IL (SSG, Gunter AFB)
    - 120 systems, 2000 interfaces (30-40% of all code)
    - Data standardization (3 ILM systems) cost $40M, 4 years
  - 7th AF (Osan, Korea)
    - TBMCS support to Integrated Tasking Order (ITO) Preparation
    - Facility target datasets failed to load (over 8,800 discrepancies)
    - ITO delivered later than required
    - Development of local work-arounds - Separate "off-line" database for aimpoints
- Network Centric Warfare (NCW) is the military's driving enabler for future operations (with information superiority, decision superiority, etc.)
  - networking the force (warfighters, weapons and C2/IT systems) for improved situation awareness, unified understanding of and action on commander's intent, etc.
    - includes Army's Future Combat System (FCS), USAF JBI, etc.
  - leverage information technology advances across physical, knowledge, and cognitive domains
  - enabled by connectivity (via Global Information Grid), "infostructures", and services
  - key enablers (from NCW DOD Report to Congress, July 2001) include:
    - connectivity, technical interoperability, sense making (semantic interoperability), integrated processes, integrated production, network-ready battlespace enablers

The Interoperability/Integration Problem (2)

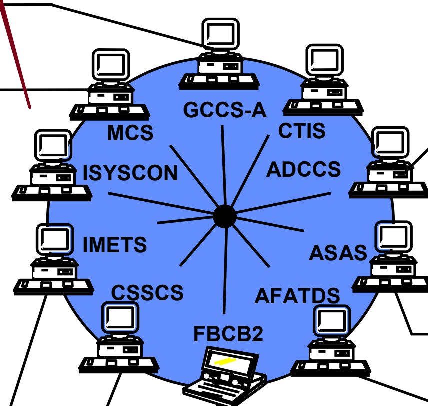
- Problem not limited to military
- B2B – "B2B today is in crisis", Dr. Marty Tennenbaum, CommerceNet (Intelink 2003)
  - billions invested, little ROI due to conceptual and structural problems
  - takes $100K and 100 days to enable a new B2B connection between 2 enterprises
  - requires new approach = "Business Service Networks": services from multiple companies within an industry loosely coupled at the *process* (vs. interface) level
- Imperfect interoperability costs $1B per year (conservative estimate) for US automotive supply chain (NIST Study, 2001)
  - mostly in repairing or re-entering data files not usable by downstream apps
- Integration is expensive
  - Glue code costs 3x more per line than non-glue code (NSF CeBASE study of COTS-Based Systems)

IXO I2 Study (Gunning, 2002): How America goes to war - Army XXI (improvement solution!)

(Courtesy of Bill Jeffrey, ATO)

SOLDIERS Are Our Credentials!

XXI



GCCS-A Interop.
- AIBS-army-96
- AMSAA-joint-96
- APC-joint-96
- ASAS-joint-97
- ATCOM-joint-96
- AWDS-joint-97
- CASCOM-joint-96
- CTAPS/TBMCS-joint-97
- DAMO-ODR-joint-96
- DES-joint-96
- DLA/ICIS-joint-96
- GCCS/GSORTS-joint-98
- GCCS/GSRDI-joint-98
- GCCS/JOPES-joint-98
- IDM-joint-98
- IOC-joint-96
- ISC-P-joint-96
- JTAV-joint-97
- LOGSA-joint-96
- MCS-joint-96
- PERSCOM-joint-96
- RAMS-joint-96
- REQVAL-joint-96
- SAMAS-joint-96
- TAV-joint-96
- Korea-allied-
- U.N.-allied-

MCS Interop.
- *LFCS-army-97
- CTAPS-joint-97
- JMCIS-joint-98
- JStarsCGS-joint-98
- TCO-joint-98
- *AUSTACCS-allied-98
- *HEROS-allied-96
- *LFCCIS-allied-97
- *E-IARRCIS-allied-98
- *SIACCON-allied-98
- *SICF-allied-96

*Part of C2SIP

IMETS Interop.
- GPS-army-
- MMS-army-
- UAV-army-
- JSTARS CGS-joint-

CSSCS Interop.
- GCCS/A-army-99
- SAMS/2-army-97
- SARSS/1-army-97
- SARSS/2A/D-army-98
- SIDPERS/2.75-army-97
- SIDPERS-3-army-98
- SPBS/R-army-97
- TAMMIS-army-99
- ULLS/S4-army-99

FBCB2 Interop.
- BCIS-army-U
- KIOWA-army-U
- LVRS-army-U
- M1A2 SEP-army-U
- M2A2-army-U
- MICAD-U
- NBCRS-army-U
- Paladin-AFCS
- PLGR-army-97
- LandWarrior-army-U
- AF (SADL)-joint-
- MC (JVMF)-joint-
- GE-allied-

ADCCS Interop.
- GBS Radar-army-94
- HIMAD-army-94
- LSDIS Radar-army-97
- TIBS-army-U
- AWACS-joint-93
- Hawkeye-joint-93
- LLAPI-allied-95
- PATRIOT-army-99
- THAAD-army-99
- JDP-joint-99
- GESAMOC-allied-99

AFATDS Interop.
- ATHS-army-97
- BCS-army-97
- FDS-army-97
- Firefinder--army-97
- FIST DMD-army-97
- FOCC-army-97
- FOS-army-98
- IDM-army-99
- IFSAS-army-97
- LtacFire-army-97
- MBC-army-97
- MDS-army-97
- MFCS-army-99
- MMS-army-97
- UAV/TS/00-army-U
- AFATDS-joint-99 (MC)
- CTAPS/TBMCS-joint-98
- IFSAS-joint-97 (MC)
- JStars/CGS-joint-97
- TacFire-01-joint-U (MC)
- ADLER-allied-98

ASAS Interop.
- Adv.Quickfix-army-U
- AMS-army-00
- ARL-army-U
- ATCAE-army-97
- CTS/CTAPS-army-97
- DAI-army-97
- Enhan.Trackwolf-army-U
- EPDS-army-97
- ETRAC-army-U
- ETUT-army-97
- GBCS-army-U
- Guardrail-army-97
- IEWCS-army-U
- IPF-army-97
- MIES-army-97
- MITT-army-97
- NGIC-army-U
- NPIC-army-U
- NSA-army-97
- SSP/S-army-97
- TEAMMATE-army-97
- TES-army-97
- THMT-army-97
- TrafficJam-army-97
- Trailblazer-army-97
- TRRIP-army-U
- UAV CGS-army-97
- UAV MPCS-army-97
- CARS/TRIGS-joint-97
- CIS- joint-
- CTAP-joint-
- IAS-joint-97
- JDISS-joint-
- JMCIS-joint-97
- JStarsCGS-joint-97
- NIPS-joint-97

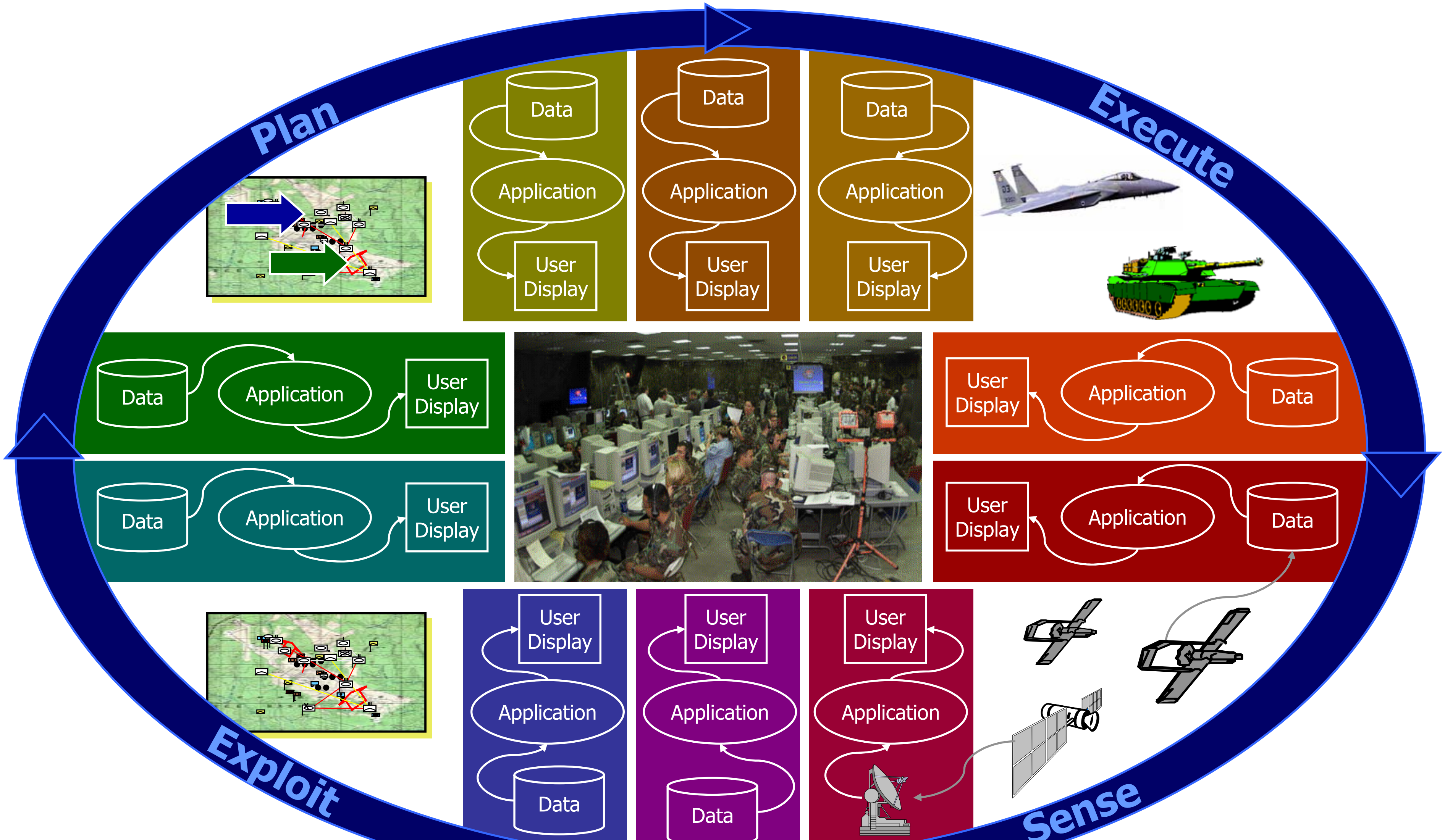
Legend:

example: *JMCIS-joint-98*

System; Interface Level: ALLIED, ARMY, JOINT; Implementation Year (U= Unscheduled)

With a services-oriented approach, systems might be broken into dozens of services. Will that help or hinder interoperability?

The Information Integration Problem



"Swivel chair" integration – slow, costly, and error prone.

Operational Example from US Army Communications-Electronics Command (CECOM)

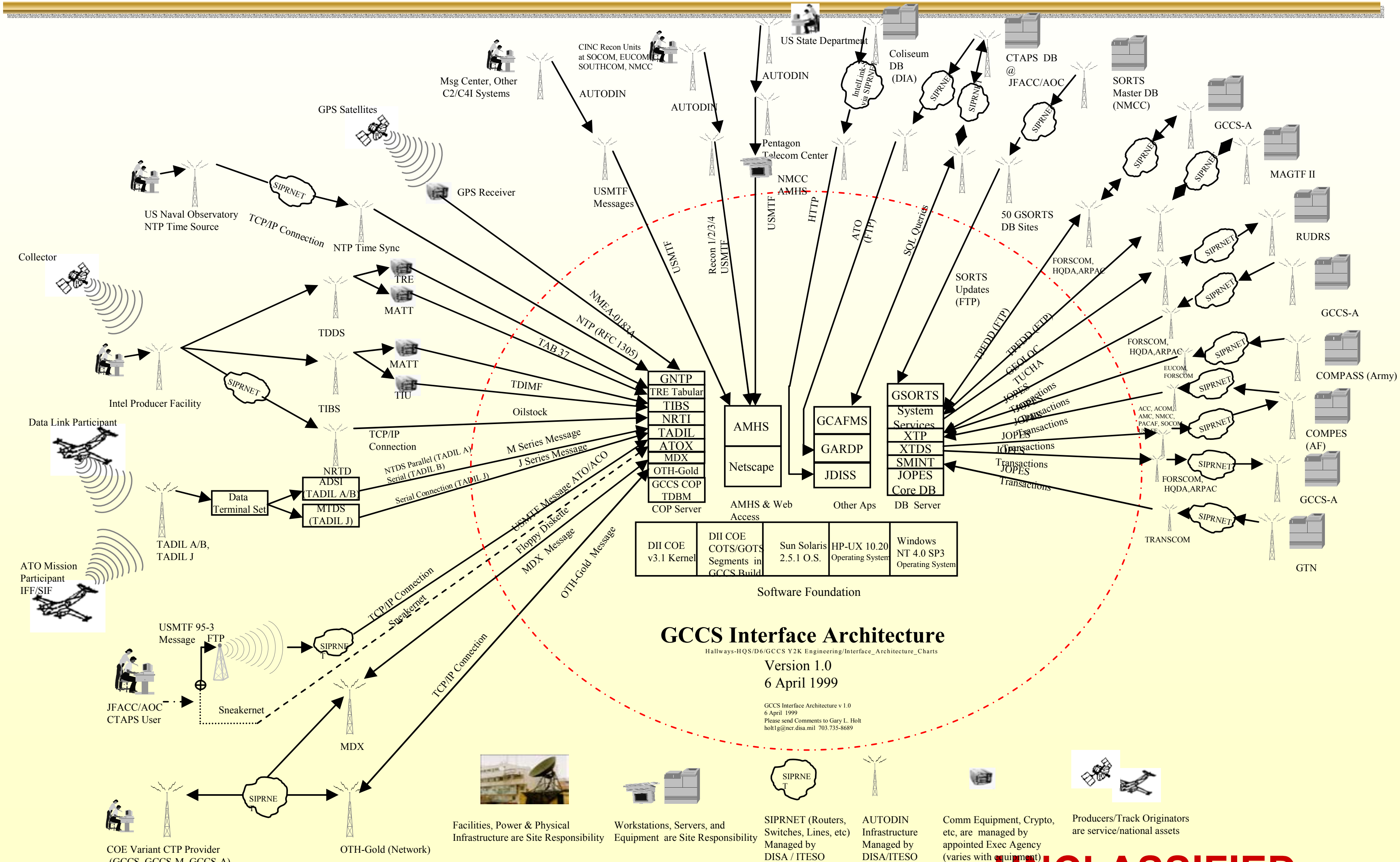
- The Problem:
  - For Operation Iraqi Freedom (OIF), information requirements of Coalition Force Land Component Commander (CFLCC) not satisfied by current systems / processes.
    - 12 critical workflows identified to share critical information across force components (e.g., USA and USMC), echelons (e.g., Corps and Division) functional areas (e.g., Ops and Intel)
    - involving 36 C2 sytems (Army with some USMC, USAF) – 200 interfaces among those systems
- Solution
  - 6 month "crash" integration effort by CTSF (Ft. Hood) & CECOM
    - required extensive coding, testing, and repair
- System replication at various echelons
  - Custom connections between some systems
  - Swivel chair integration between some systems
- Impact:
  - Resources consumed (programmers, hardware, training, etc.)
  - **Implementation of these workflows may have contributed to the delayed start of OIF**
  - **C2 interoperability issues hampered flexible use of forces**

Sample Thread

- Blue Situation Awareness (SA) information sharing
  - across echelons: e.g., between Corps and Division level
    - share blue position information
    - between GCCS-A and MCS-L systems
  - between Operations and Intelligence
    - between GCCS-A/C2PC and ASAS systems
    - integrate display of Red and Blue situation information to find threats to Blue
- Challenges
  - location of systems – some replication of hardware/software required (plus training)
  - limited configuration possible (e.g., GCCS output message modes)
  - interoperability incompatibilities between original workflows
    - e.g., reference data incompatibilities for Unit ID Codes (UICs)
  - documentation not always available and consistent
    - e.g., System of Systems manual has discrepancies in ???% of entries
  - programmers required
  - human still in the loop as "translator" (swivel chair interoperability)

UNCLASSIFIED

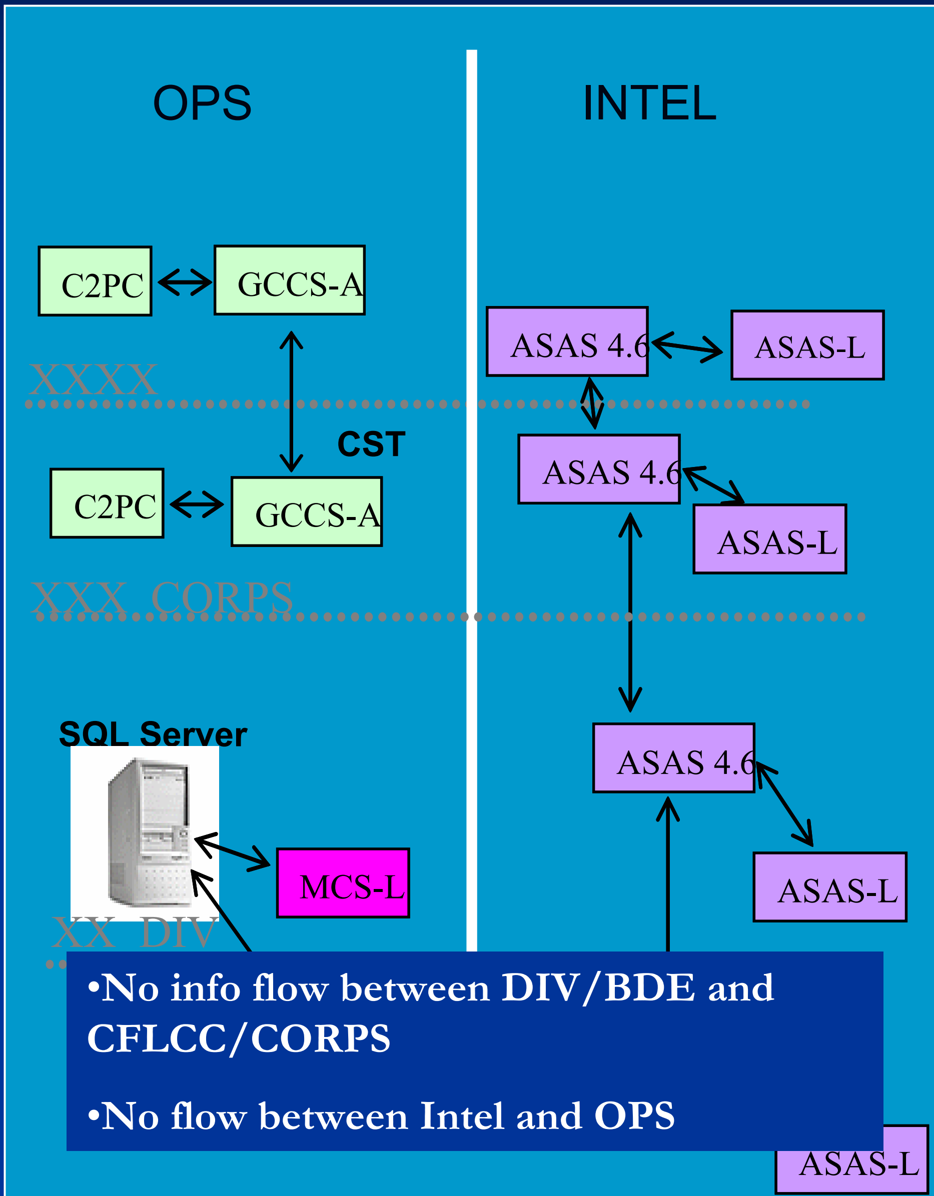
GCCS Interface Architecture



UNCLASSIFIED

CTSF/OIF Blue SA Thread: Architectures

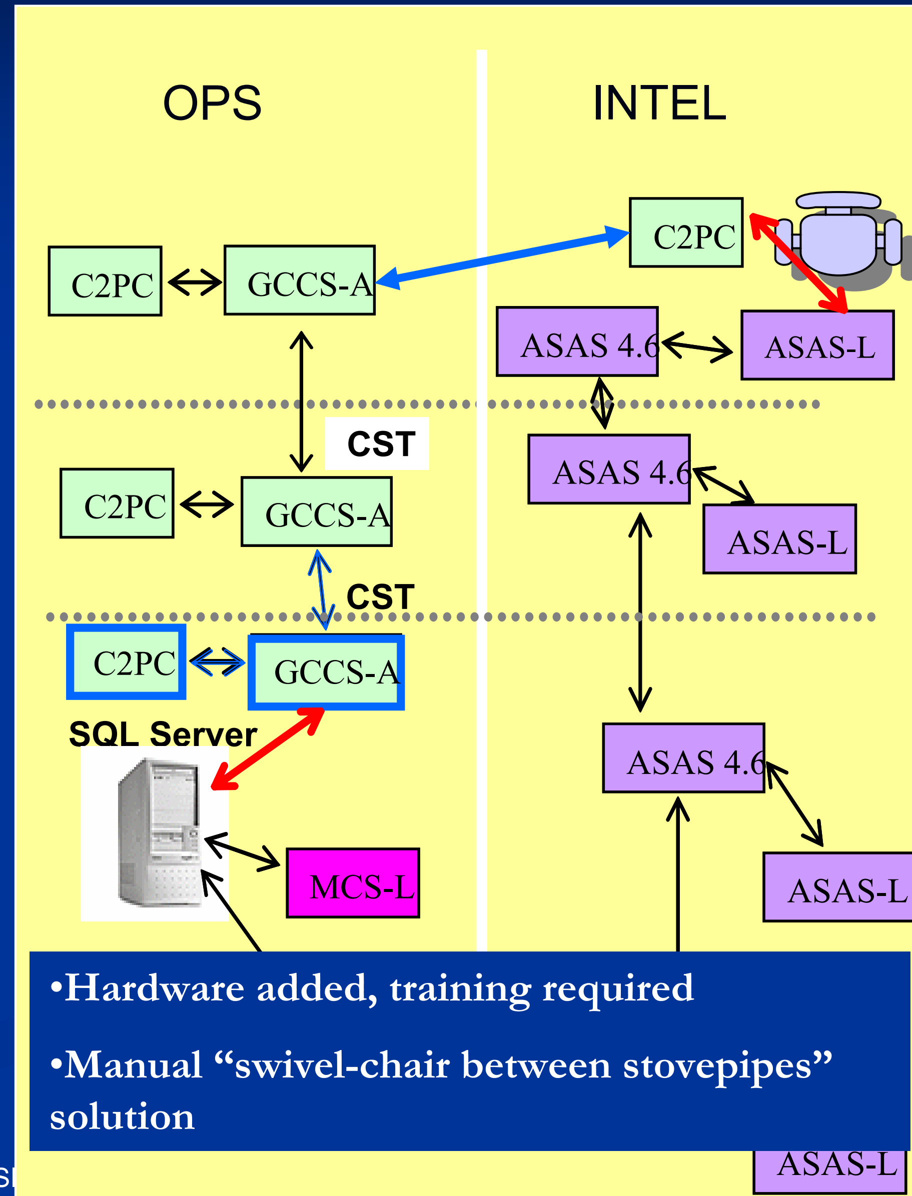
Original Architecture



•No info flow between DIV/BDE and CFLCC/CORPS

•No flow between Intel and OPS

Improvised Architecture



•Hardware added, training required

•Manual "swivel-chair between stovepipes" solution

Future Concept of Operations: Assured, Improvisational Interoperability

- Support improvisational workflows through rapid generation of a custom system of systems by non-programmers
  - new workflow identified to support a commander's information/C2 need and specified by a non-programmer
  - automated assembly of (wrapped) component systems (and services)
  - ad-hoc interoperability (at the process/operational level)
    - may be custom "one of" system of systems
  - assured interoperability
    - analysis (V&V, etc.) of system of systems for correctness, completeness, quality of service, etc. (prior to and during execution)
- Move from data-level interoperability to process-level interoperabiliy
  - requires semantics
- DoD is moving towards services-oriented architecture, along with the commercial world
  - How can this be exploited to revolutionize military command and control?

Assured Interoperability: Risk and Sources of Error

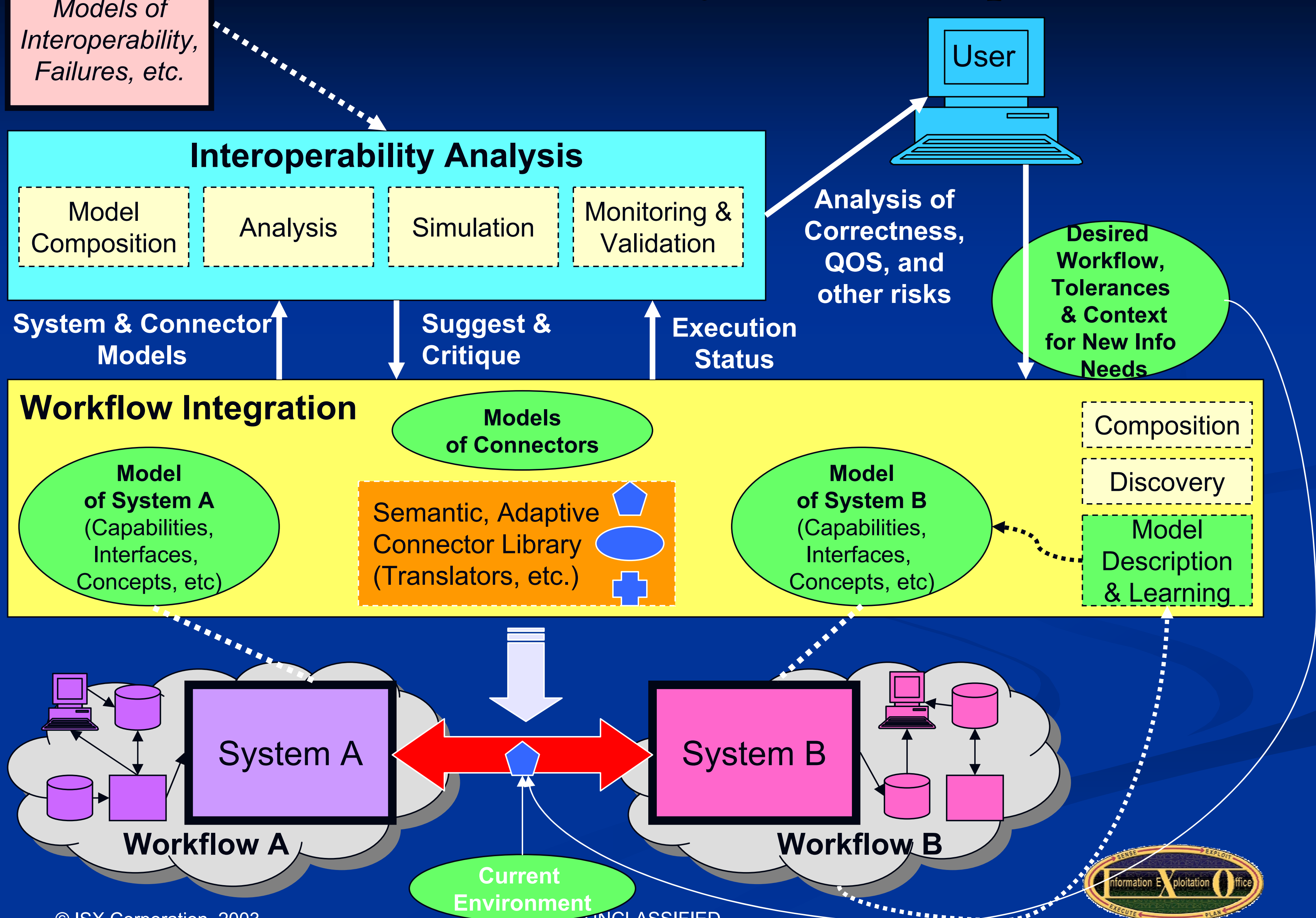
- Differing Syntax – may cause parse failure
- Differing Terminology
  - different units – yards vs. feet
  - different names for the same class (e.g. Employee vs. Worker)
  - different names for same entity (including abbreviations)
    - e,g, Mass. vs. Massachusetts vs. The Bay State
    - e.g., plane observed at Airfield X vs. dark shadow in satellite imagery photo 103
- Differing Concepts/Ontologies
  - differing coordinate schemes
    - origin point
    - dimensionality
  - different concepts with the same name
    - may differ in granularity – e.g., 1776 vs. 18th Century; Paris vs. France vs. Europe; engine vs. entire car
    - may be related by subsumption
  - different abstraction hierarchies (class hierarchies)
- Differing Values for the Same Attribute (Data Discrepancies)
  - different values
  - same value but at differing precision/resolution
- Different Reference Data
  - different sources (check information pedigree)
  - different accuracy/precision
- Different Context
  - different mission objectives
  - different assumptions or constraints (ROEs, etc.)
  - different views of the battlespace
- Different Workflows
  - different target workflow
  - different original workflow
- Different Timing
  - different synchronization, latencies
  - different updates
  - different resource utilization (can lead to deadlocks, etc.)

Example: 2 route planners:
•different inputs – origin and destination: coordinate schemes
•different maps – from different GIS sources, of different scales, different versions of same map
•different outputs – waypoints vs. line segments; coordinate schemes; scale
•different models and methodologies – route planning algorithms, doctrine/ROEs/threat models/vehicle capability models (can't cross X, go near Y), etc.

Some Requirements

- Explicit semantic representations
  - for models of system interfaces, processes, workflows, etc.
  - wrapping of systems, components, and services (by programmers, with automation)
- User interfaces
  - GUIs for non-programmers to sketch new workflows
  - GUIs to display results of new workflows
- Automated assembly of systems, components, and services
  - discovery
  - composition using adaptive, semantic connectors
- Automated analysis of composed system of systems
  - compose models of components
  - analyze composed model for correctness, completeness, QOS

A Notional System Concept



Making the Case

- Some Questions
  - What is the operational need?
  - Why DARPA? Why Now (versus yesterday, tomorrow)?
    - why this is distinct from – yet builds on – previous DARPA work (I3, HPKB, RKF, ARPI/Planning, CoABS, DAML, etc.)
  - If DARPA invests $*n* million in this area, what is the argument it will (1) be successful and (2) produce a good ROI
- Make a quantitative case where possible – some factors:
  - processes (old and new) being automated
    - expected level of automation (via technologies x, y, z…)
    - projected benefits over the next several years
  - expected operational impact
    - better, cheaper, faster, etc. – specify metrics to evaluate
  - e.g., incremental cost of adding semantics, etc. on top of DOD SOA(s)
- Use operational examples to ground the model
  - challenge is to get the data
  - argue from these "base cases" by induction

Influence Diagram Model

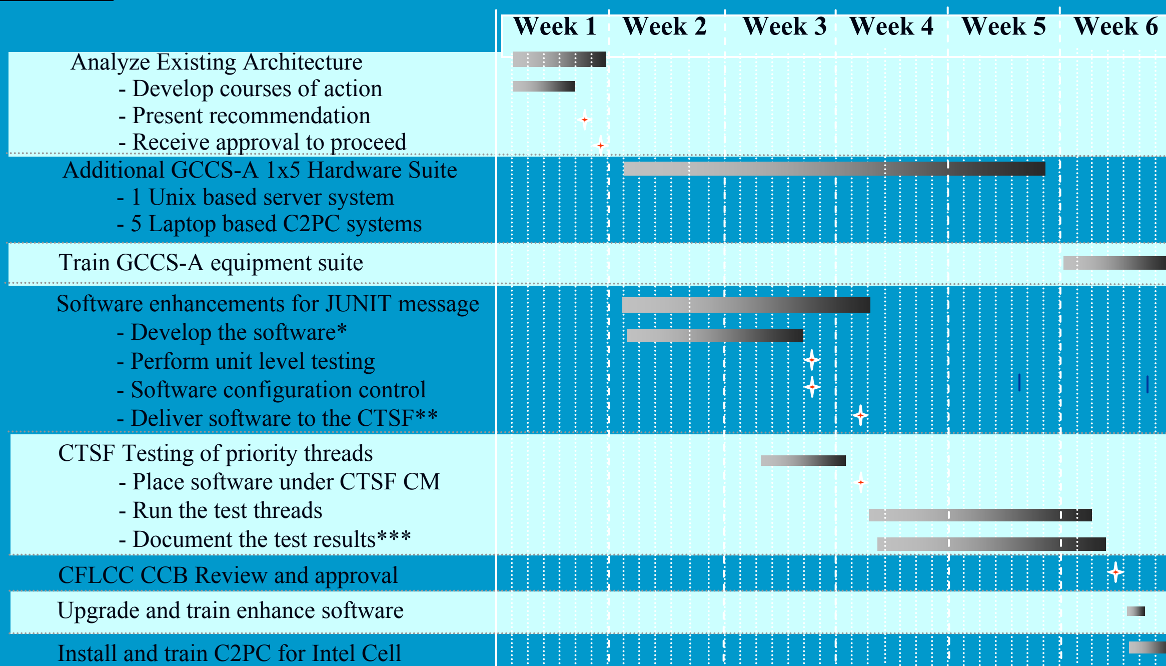
- Why influence diagrams?
  - Captures relationships ("influences") quantitatively
  - Graphical representation (versus buried in a spreadsheet)
  - Allows easy "what if" and sensitivity analysis
  - Easy to tweak – e.g., change values/distributions of input parameters, functions, etc.
- Used commercial influence diagram tool: Analytica by Lumina Decision Systems
  - can generate data for graphing via Excel

Model Rationale

- Model the integration of two systems
- Key metrics
  - time
    - does <u>not</u> include time to build connectors
    - does include selection, configuration, and application of connectors
    - does include time to model systems
  - correctness
    - error rates from various steps
    - does not yet model cascading errors
  - completeness
    - still fuzzy
      - includes schema elements mapping "recall"
      - could also include % of data translated by semantic connectors at runtime
    - exception handling ability
      - majority of workflow definition effort
      - analogy to automated test case generation
  - QoS attributes – not yet modeled
  - Comparison – non-SEE, SEE Baseline (current tools), SEE Level I and II
    - SEE Baseline might also include WSDL-ified services (e.g., DISA NCES)
- Inputs
  - Problem attributes – size and complexity of integration (# of schema elements, original workflows), "semantic distance" between the 2 systems
  - Skills – modeling, etc. (programmer and domain skills)
  - Technology contributions (done as % improvement over baseline)
- Constants, etc. validated from CECOM data and from literature survey of relevant areas: e.g., COCOMO II for software engineering metrics (non-SEE case)

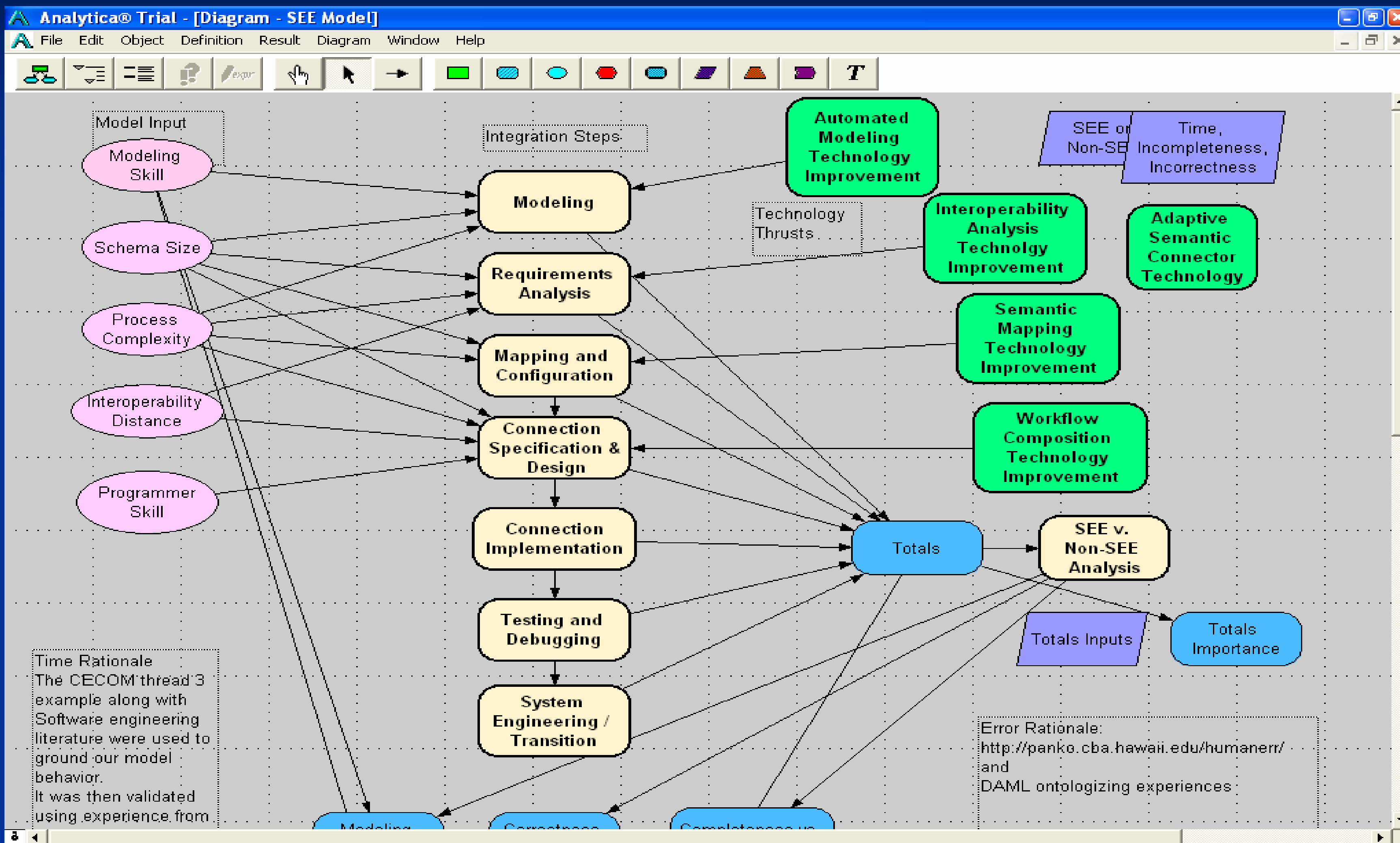
CECOM Example – Thread 3

Timeline



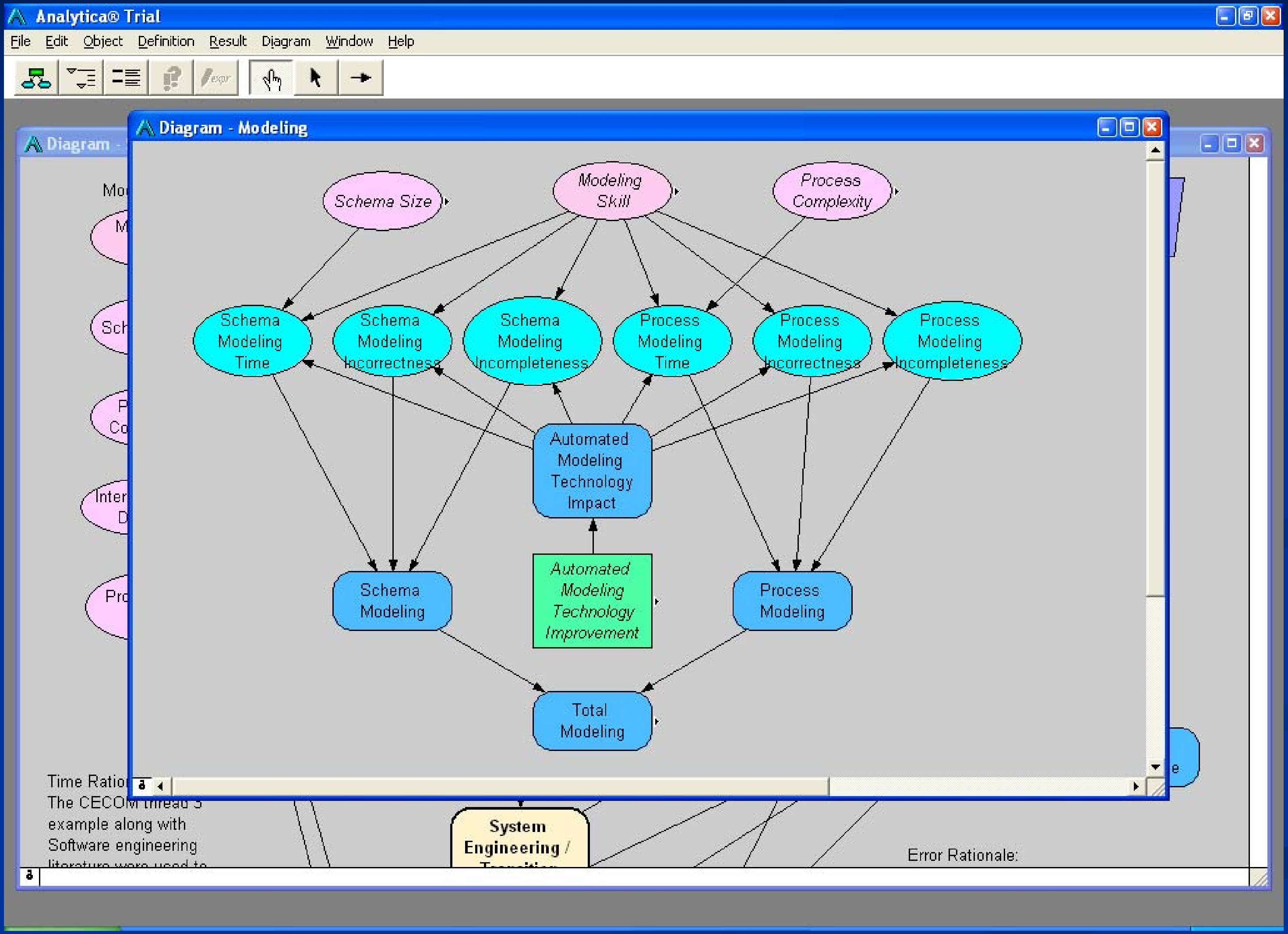
*Software developed by one senior and one junior programmers
**Software delivery occurs by ftp download
***Test results documented at the end of each day, so a draft report was available at the end of testing

SEE Influence Diagram



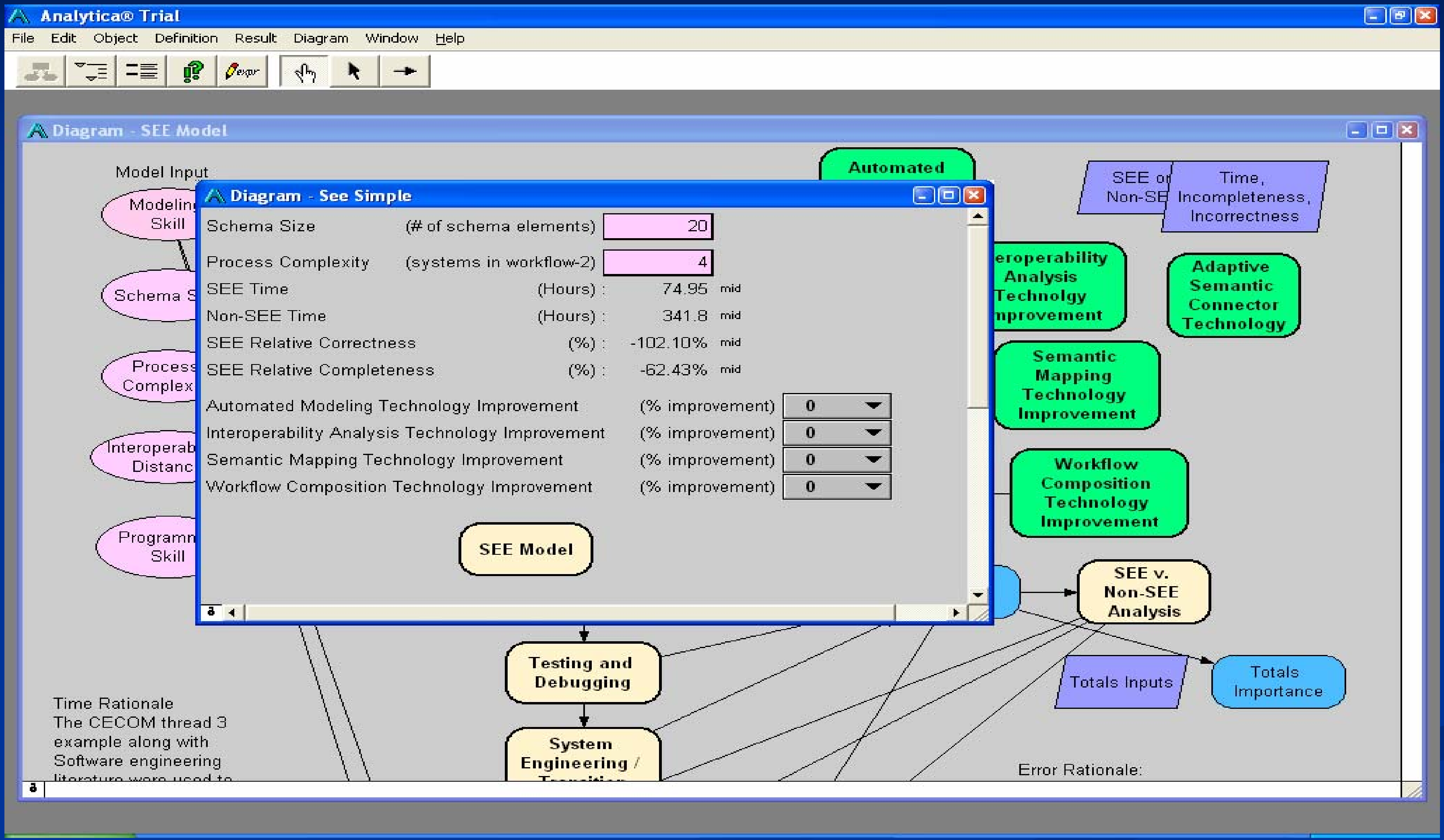
UNCLASSIFIED

Modeling Detail



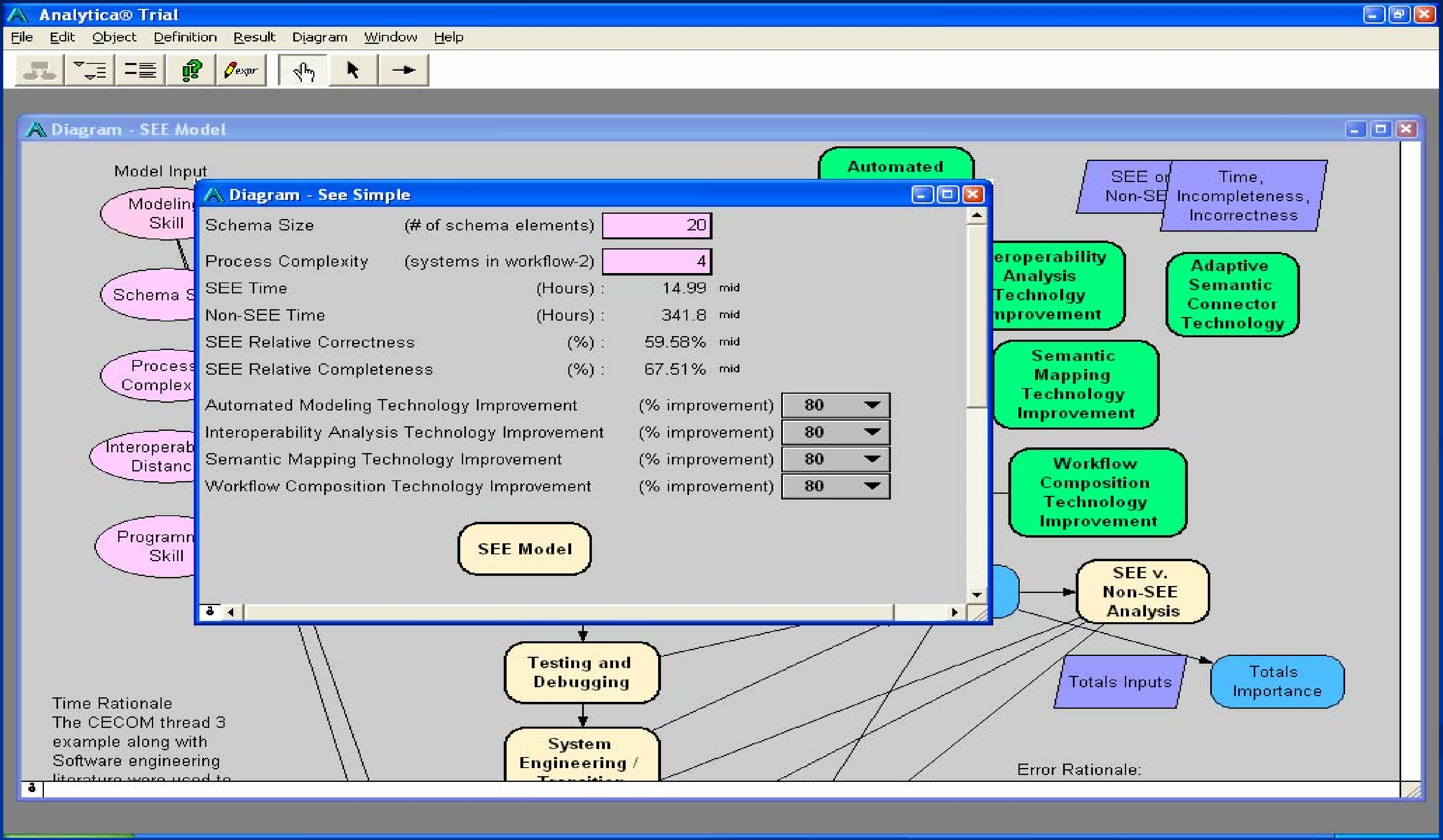
UNCLASSIFIED

Thread 3 Results

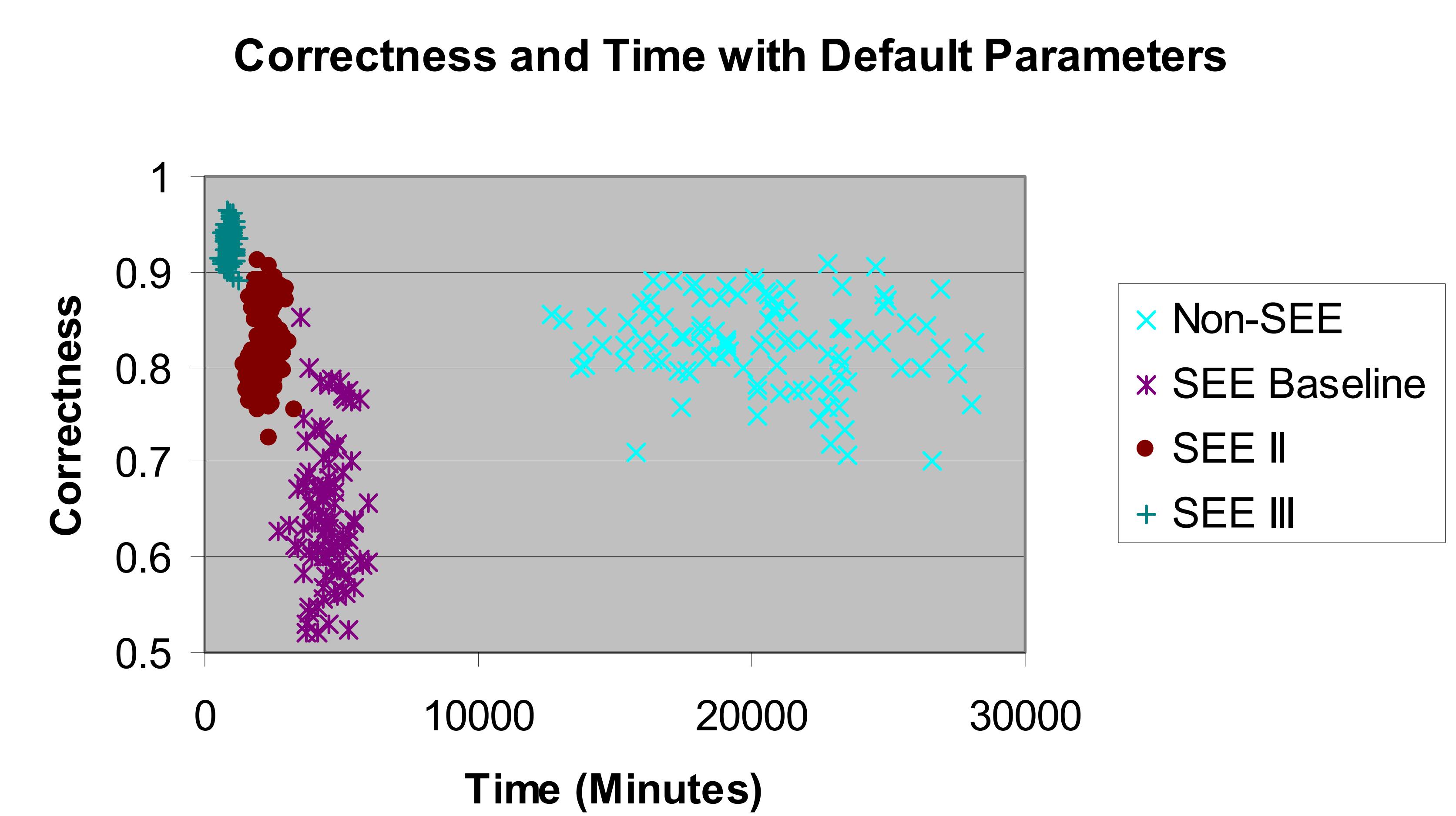


UNCLASSIFIED

Thread 3 with Tech Improvement



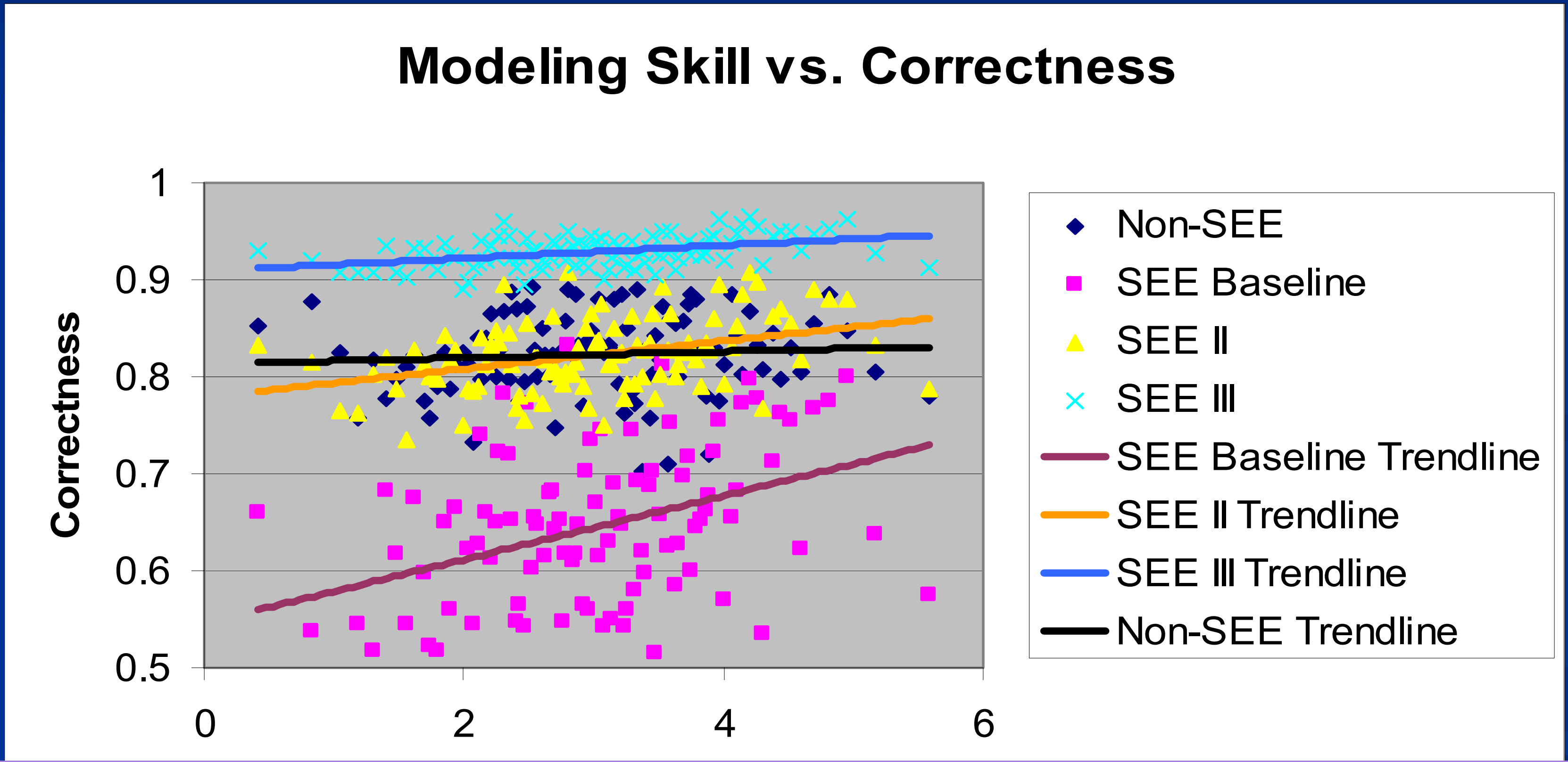
Model Results: Correctness vs. Time



SEE Versions II and III (Baseline + 50% & 80% on all 4 technology area) reduces time and improves correctness (and variability).

Assumptions: CECOM Example schema/workflow size. Independent Variables Time, Correctness

Model Results: Modeling Skill vs. Correctness



SEE Versions II and III (Baseline + 50% & 80% on all 4 technology area) reduces errors (across all skill levels) and the impact of skill in general.

Technology Survey Approach (1)

- Determine key technologies (decompose the "green boxes")
  - current state-of-the-art
    - commercial world, research community
  - projected capabilities over next several years (preferably quantitative)
    - commercial world, research community
  - expected contributions to operational goals (metrics)
  - literature survey and discussions with leaders in field
    - areas: workflow systems, interoperability, semantic reasoning/web, software agents, software systems analysis, interoperability infrastructures, etc.
- Does SEE require
  - (1) breakthrough(s) in one or more key technologies
  - and/or, (2) unique "recipe" to combine one or more key technologies?
- Why DARPA?
  - Does military have unique needs that commercial is not addressing (and won't be for next several years)?
  - Does technology require a significant breakthrough on the research front?

Technology Survey Approach (2)

- Work should mine previous/current/future DOD investments
  - e.g., DARPA I3, HPKB/RKF, CoABS, DAML, etc.
  - DISA NCES, USAF Joint Battlespace Infosphere
- Work should complement standards processes
  - improvisational workflows require rapid interoperability
    - dynamic exchange of content, negotiation of protocols, etc.
  - can inform standards processes: e.g., standards cycle is a slower moving "outer loop" (avoid one size fits all, etc.)

Some Key Technical Challenges

- Modeling systems, processes, and workflows
  - representations and formalisms that go beyond interfaces
  - semantic grounding (vs. WSDL)
  - recovering models from legacy systems
- Connecting systems
  - types of connectors and applicability
  - representation of context
  - automated adaptation of connectors based on context/environment
- Analyzing interoperability
  - composing models of systems and reconciling diverse semantics
  - reconciling diverse semantics
  - using models to predict correctness, QOS, etc.
- Usable user interfaces for
  - specify new information requirements and sketch new workflows
  - understand risks
  - model new systems

Applicable Technologies – Modeling of Systems' Interfaces, Schemas, Workflows

- Support a user (programmer or domain expert) in creating machine-understandable models for systems' interfaces, schemas/concepts, and workflows)
  - semantic representations (e.g., ontologies/OWL)
  - web services (e.g., XML-based – WSDL, XPDL, BPEL4WS)
  - semantic web services (e.g., DAML-S, OWL-S)
  - other process representations – e.g., PSL, Petri Nets, Process Algebras, CSP, CCP, etc.
  - automated modeling tools to generate models for legacy systems/services
    - automated ontology generation
    - machine learning of schemas
  - model validation

Applicable Technologies – Workflow Integration

- Support a user (nonprogrammer) in assembling a system of systems for an improvisational workflow
  - Discovery
    - e.g., directories (UDDI, etc.), matchmakers
  - Composition of systems/services
    - semi-automated – e.g., U.Md. Web Services Composer
    - automated – e.g., planning for (semantic) web service assembly
    - COTS workflow engines and web services orchestration engines
  - Connectors to link component systems/services (with configuration) and adapt to changes during execution
    - Translators – e.g., XSLT, schema/ontology mapping
    - Intelligent Agents for execution monitoring, repair
  - Integration infrastructures – e.g., .NET, DISA NCES, J2EE, Jini, CoABS Agent Grid, other agent frameworks
  - HCI
    - specify new workflow – e.g., GUIs
    - visualize information output from new workflow
    - visualize workflow properties (e.g., correctness analysis, risks) and execution status

Applicable Technologies – Interoperability Analysis

- Support a user (nonprogrammer) in V&V of a system of system (prior to and throughout execution)
  - Model composition
  - Model analysis (for correctness, QOS, etc.)
    - various techniques depending on modeling formalism(s) used – e.g., safety properties for Petri nets, theorem proving, etc.
  - Simulation

Help Wanted

- (More) operational examples and analysis
  - inputs to model from real-world studies
- Refine key technologies list
  - assess state of the art
    - today and tomorrow
    - commercial world and research community
- Help formulate program/research agenda
- Id transition partners
- Make the case – weave above into a sellable story
  - 25,000 foot view / elevator pitch
  - detail for drilldown

Questions?

- Acknowledgements
  - DARPA
    - Dr. Robert Tenney and Dr. Mark Greaves
  - US Army CECOM
    - Mr. John Soos and Mr. Jeff Harrold
- Email:
  - Mark Greaves (mgreaves@darpa.mil)
  - Brian Kettler (bkettler@isx.com)